CLAIMS

We claim:

1. A computer-implemented method for filtering messages routed across a network, the messages including field name-value pairs, the method comprising:

extracting field name-value pairs from the messages;

determining, for values of the same field name, a most restrictive data type of the values; and

storing the data type in association with the field name.

2. The method of claim 1, further comprising:

generating a rule which would allow messages having values of a field name that match the most restrictive data type.

3. The method of claim 2, further comprising:

applying the rule to determine whether to allow messages having values for a field name that match the most restrictive data type.

4. The method of claim 1, wherein the determining step further comprises:

determining a match factor for a data type, the match factor indicating a fraction of values for the same field name that match the data type; and

selecting a data type having a match factor exceeding a threshold and having no child data types with a match factor exceeding the threshold.

5. The method of claim 4, wherein the threshold is a fraction of values for the same field name which should match the data type.

6. A computer program product comprising a computer-readable medium containing computer program code for performing any one of the methods of claims 1 through 5.

7. A computer-implemented method for filtering URL messages routed across a network, wherein the messages include URL components, the method comprising:

extracting URL components from the messages;

determining, for URL components at the same level, with the same root URL component, a most restrictive data type of the URL component; and

storing the data type in association with the URL component.

8. The method of claim 7, further comprising:

generating a rule which would allow messages having the URL components that match the most restrictive data type.

9. The method of claim 8, further comprising:

applying the rule to determine whether to allow messages having the URL components that match the most restrictive data type.

10. The method of claim 7, wherein the determining step further comprises:

determining a match factor for a data type, the match factor indicating a fraction of URL components at the same level, with the same root URL component, that match the data type; and

selecting a data type having a match factor exceeding a threshold and having no child data types with a match factor exceeding the threshold.

11. The method of claim 10, wherein the threshold is a fraction of URL components at the same level, with the same root URL component, which should match the data type.

12. A computer program product comprising a computer-readable medium containing computer program code for performing any one of the methods of claims 7 through 11.

13. A computer-implemented method for inferencing a data type of scalar objects, the method comprising:

determining a match factor for a data type, the match factor indicating a fraction of scalar objects that match the data type; and

selecting a most restrictive data type having a match factor exceeding a threshold and having no child data types with a match factor exceeding the threshold.

14. The method of claim 13, wherein the threshold is a fraction of scalar objects which should match the data type.

15. A system for inferencing a data type of scalar objects, the system comprising:

a module for determining a match factor for a data type, the match factor indicating a fraction of scalar objects that match the data type; and

a module for selecting a most restrictive data type having a match factor exceeding a threshold and having no child data types with a match factor exceeding the threshold.

16. A computer program product comprising a computer-readable medium containing computer program code for performing any one of the methods of claims 13 through 14.

17. A system for filtering messages routed across a network, the messages including field name-value pairs, the system comprising:

a learning engine, for extracting field name-value pairs from the messages, determining, for values of the same field name, a most restrictive data type of the values, and storing the data type in association with the field name; and

a message filter, for generating a rule which would allow messages having values of a field name that match the most restrictive data type.

18. The system of claim 17, wherein the learning engine is further adapted to generate a rule which would allow messages having values of a field name that match the most restrictive data type.

19. The system of claim 17, wherein the message filter is further adapted to apply the rule to determine whether to allow messages having values for a field name that match the most restrictive data type.

20. The system of claim 17, wherein the learning engine is further adapted to:

determine a match factor for a data type, the match factor indicating a fraction of values for the same field name that match the data type; and

select a data type having a match factor exceeding a threshold and having no child data types with a match factor exceeding the threshold.

21. The system of claim 20, wherein the threshold is a fraction of values for the same field name which should match the data type.

22. A system for filtering URL messages routed across a network, wherein the messages include URL components, the system comprising:

a learning engine, for extracting URL components from the messages, determining, for URL components at the same level, with the same root URL component,

a most restrictive data type of the URL component, and storing the data type in association with the URL component; and

a message filter, for generating a rule which would allow messages having the URL components that match the most restrictive data type.

23. The system of claim 22, wherein the learning engine is further adapted to generate a rule which would allow messages having the URL components that match the most restrictive data type.

24. The system of claim 22, wherein the message filter is further adapted to apply the rule to determine whether to allow messages having the URL components that match the most restrictive data type.

25. The system of claim 22, wherein the learning engine is further adapted to:

determine a match factor for a data type, the match factor indicating a fraction of URL components at the same level, with the same root URL component, that match the data type; and

select a data type having a match factor exceeding a threshold and having no child data types with a match factor exceeding the threshold.

26. The system of claim 25, wherein the threshold is a fraction of URL components at the same level, with the same root URL component, which should match the data type.